International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 4 Issue 06, June-2015

Segmentation and Detection of Text in Natural Scene Images

Meghana Thodaskar N R
M.Tech, ISE
PESIT
Bangalore VTU INDIA

Rama Devi P
Assistant Professor
ISE PESIT
Bangalore, VTU INDIA

***Abstract*— Any text data present in natural scene images contains useful information like text based landmark etc. Extraction of text from scene images involves many stages. Each every stage is equally important to get efficient results. Detecting the text, localising the text, and segmentation, recognition are important steps. Extract.ion from scene text images is very difficult due to variations in size, orientation, alignment from one image to another. From all these difficulties extract.ion of text from scene images is a challenging task. However, text in such images is not confined to any page layout, and its location within in the image is random in nature. In addition, motion blur, non-uniform illumination, skew, occlusion and scale-based degradations increase the complexity in locating and recognizing the text in a scene/born-digital image. In our proposed method we have used The otsu binarization TECHNIQUE. It is applied on separated R, G, B channels. Based on connected component information the text is extracted. This method is implemented to handle light and night scene images.**

Keywords*— *Connected-Components, Dilation, Otsu-Method, Logical "And" Operation, Morphological, Rgb Channel

I. INTRODUCTION

In computer_vision field document analysis and recognit ion (DAR) has a great history for more than four decades. This duration it self indicates the complexity involved in the field. DAR has expanded rapidly by frequently adding new sub-fields for innovative research. This makes more interest in the process of evolution. The major difficulties found in the field are photocopying, printing, scanning and image capturing technologies. Hence the idea of completely recognizing the content of a document is amazing and is a task still difficult to achieve by a machine, unlike the human brain. There are a large number of other applications that have been developed by researchers, with various advancements in technology such as photocopying , Xerox, revolutionized the field of documents. The photocopying machine does not possess this capability. Today optical character recognition (OCR) engines reasonably perform this task by creating a soft copy of the document.

Along with so many applications the scanner has great use. It was invented to convert printed, handwritten and historical documents into digital format. This helps the conversion process in archives. Depending on the required scanning resolution a single scanned image can consumes a large storage space. There is a huge difference in ratio between storage space consumed by a scanned image and information stored as coded text. This has motivated many researchers to take up this topic for their research work. The main goal is to work on methods to recognize the textual content of a scanned image. By this pattern recognition taught the complexities to recognize Roman characters. Thus, digital documents gave birth to the analysis of actual content in the digital documents such as text or data mining, information retrieval, and relation between documents.

In today's world each and everybody has mobile phones. Due to the availability of low-cost cameras and mobiles with a camera, one can able to create a camera captured documents. The portable handheld devices overcome the limitations of a scanner and also increased the range of documents. Imaging the text on a notice board with a glass frame is an example for the limitation of a scanner. This limitation can overcome by using mobile device. The mobility of cameras makes it possible to capture the contents of a notice board from any reasonable distance. Billboards and signboards are the examples of camera captured documents. Usually a captured image has less text than scene. Sometimes text may or may not be present in many scenes. Therefore detecting the presence or absence of text in a scene image is a major problem. This process is called text detection problem. It is almost like asking a blind person to analyze the scene. To reduce this complexity the problem itself broken down into pats such as text localization and recognition

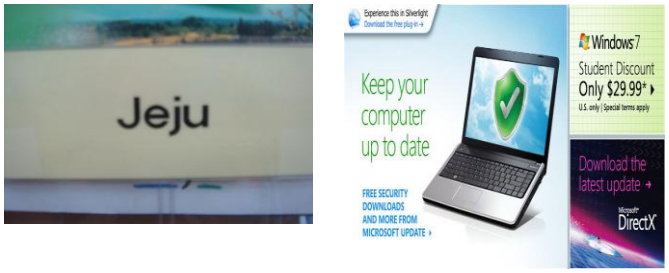


Figure 1: Sample born-digital/scene images from KAIST dataset

II. RELATED WORK

Hongliang et al [1] in this paper author describes an efficient technique of locating and extracting license plate and recognizing each segmented character [4]. The proposed model can be subdivided into four parts- Digitization of image, Edge Detection, Separation of characters and Template Matching. Morphological– operations with structuring element (SE) used to eliminate non-license plate region and enhance only the plate region. Character segmentation is done using Connected Component Analysis. Correlation based template matching technique is used for recognition of characters.

Bai et al [2] presents method for Chinese tex.t recognition in images/videos. The method is different from existing one which binarized text images, fed binarized image to an OCR and gets the recognized results. The proposed scheme implements the recognition directly on gray pixels followed by segmentation, building recognition graph, Chinese character recognition and beam search determination. The advantages lie in, it does not depend on the performance of binarization, which is not perfect in practical and thus decrease the performance of OCR and grayscale image gives more information of the text which in turn helps in improving recognition rates.

Neha et al [3] employ a dis.crete wave_let trans.form (D.W T) method to extr.act text infor.mation from complex-background. The input coul.d be a color image or a grayscale. Sobel edge detection method is used to find out edges on each sub-image. The obtained result is considered to form an edge map. In the next step morphological operations are applied on edge map and further thresholding is applied to improve performance.

Dutta et al [4] the method is based on the gradient information and edge map selection. As an initial step the algorithm first find the gradient of the image and then enhance the gradient information. In the next step binarized the enhanced gradient image and select the edges by taking the intersection of the edge map with the binary information of the enhanced gradient image. To generate the edge map canny edge detector is used. The selected edges are then morphologically dilated and opened using suitable structuring elements and used for text regions. To identify the boundary of the text regions projection profile analysis is performed.

Sivasankaran et al [5] the authors used grayscale transformation and smoothing using median filter as pre processed steps. Canny edge detection and Gaussian filter method is used to remove weak edges. Further with dilation and connected component labelling techniques text part is extracted.

Angadi et al [6] proposed work is based on tex.ture_analysis and uses discre.te co.sine trans.form (D.C.T). This uses high pass filter to remove similar background. The resultant texture_features are then applied on each 50*50 block of the input and strong text blocks are identify.ied using discrim.in-ant fu.nctions. At last, the detected text blocks are added and get the extracted text regions.

Wei et al [7] the authors used a pyramidal concept to detect_text in video im.ages with variations of background, size_of text font and colour. In the first step, t wo down_sized images are obtained from the origin_al image. Then, the gradi ent difference is calculated for three differentl y sized images. k-m.eans clustering procedu.res are applied to separate the pixels. Next, determine the boundaries of candidate text regions using projection _profile analysis. Finally, text candidates are identified using two verification phases. One is geometric_properties. Another is text candidate using D.WT. To reduce the number of dimensions of these features principal component analysis is used. SVM is used to classify the text and non-text.

III. PROPOSED METHOD

The Algorithm and flowchart of the proposed method using Otsu binarization is as shown below:

Algorithm 1

Step1: Read input image.
*Step2: Resize the input image to 530*600.*
Step3: Separate the R, G, B channel.
Step4: Apply Otsu binarization on each individual plane
Step5: Complement form is used to identify the text of inverse polarity on each binarized plane.
Step6: Perform logical "and" operation to combine step 4 and step 5.
Step7: Use morphological operations for segmentation.
Step8: Binarized the edge image enhancing only the text regions against a plain black background.
Step9: Apply bounding box to localize the text region using connected components.
Step10: Final result.

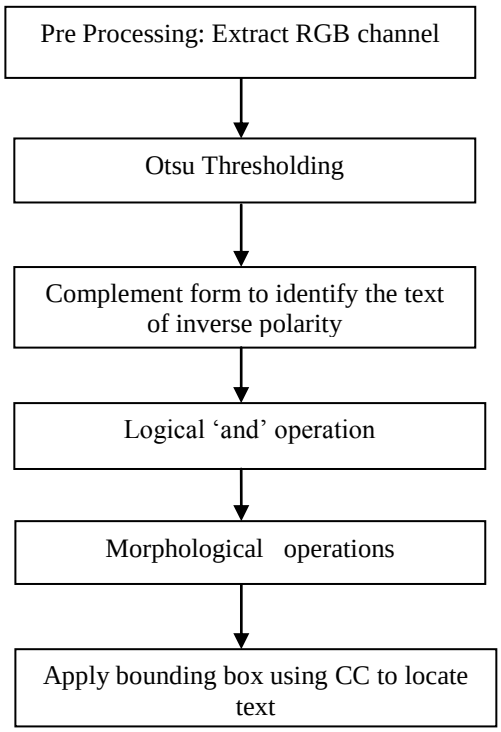


Figure 2: Flow chart of proposed method.

A. *Preprocessing*

The proposed method is based on color based character extraction. However, color information is also important because, usually, related characters in a text have almost the same color for a given instance encountered in the scene. The color of each pixel is determined by the combination of the red, green, and blue intensities stored in each color plane at the pixel's location. In the RGB model, an image consists of three independent image planes, one in each of the primary colors: red, green and blue. The three (Red, Green and Blue) plane pixel values were used as feature vectors.

Algorithm 2

Step1: Read input image.
*Step2: Resize input size to 550*600.*
Step3: separate R,G,B channels.

The extracted R,G,B channel is as shown in the figure 2. We apply otsu binarization technique on each extracted R,G,B plane.We will explain this step in detail in section B.

2(a)

2(b)

2(c)

Figure 3: (a)(b)(c) shows extracted R, G, B, channels

B. *Otsu Thresholding*

Each plane is separately segmented using Otsu Global Thresholding (ogt). Otsu binarization is an effective method for segmentation when the variations in lighting and colour are minimal. The histogram of an image is used to arrive at the threshold that maximizes the discrimination value. The values in the histogram are normalized before calculating the discrimination value. At each gray value, the histogram is split into two parts. The mean and weight of each histogram part are calculated, and also the discrimination value. The gray value at which a peak is found for the discrimination value is used as the global threshold. The sum of vari-ance is calculated using the below formula:

$$\sigma^2 \omega(t) = \omega_1(t)\,\sigma^2{}_1(t) + \omega_2(t)\,\sigma^2{}_2(t) \qquad (1)$$

The Results of Otsu's binarization of colour channels is as shown below:

Redmi 4(b) 4(c)

Figure 4: Results of Otsu's binarization of colour channels for one of the sample image in Figure1 (a) Binarized red plane. (b) Binarized green plane (c) Binarized blue plane.

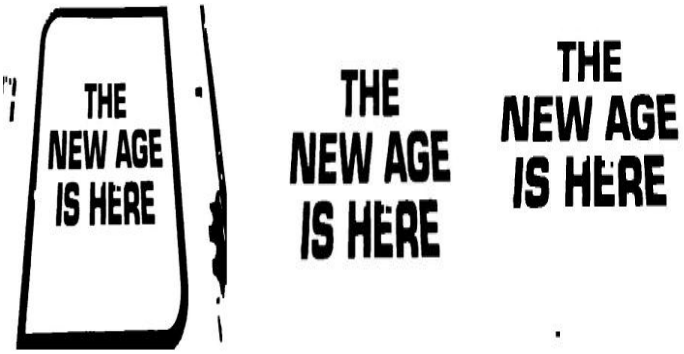
We observed that applying Otsu separately on RGB components often recover lost texts efficiently. Otsu's method as described above converts several pixels from foreground to background and also vice versa

C. Complement form to identify the text of inverse

Apply morphological open operation on each binarized plane and take complement of these images to eliminate the noises of an image. Complement version of the results of the figure 5 is shown in figure 6. It is observed that part of the text that was merged with the background becomes foreground in the complement image and can be successfully captured. In the complement of an RGB image, each pixel value is sub-tracted from the maxi-mum pixel value and the difference is used as the pixel value in the out_put im.age. In the output image, dark areas become lighter and vice-versa.

5(a) 5(b) 5(c)

Figure 5: (a) (b) (c) Dilation on binarized red, green, blue channel. It can be seen that the text that is lost in binarization in one color plane is captured in some other color plane



6(a) 6(b) 6(c)

Figure 6: (a)(b)(c) Complement version of the results of the Figure 5(a)(b)(c). It is observed that part of the text that was merged with the background becomes foreground in the complement image and can be successfully captured

D. Logical operations

As a next step we apply and operation. In our proposed method we take combination of each binarized plane and its complement version. This step refers to identify the characters as they are in original image. This is done by multiplying resultant image (figure: 5) with binary converted complement image (figure: 6). In this method, pixels having value 1 represents text and pixels having value 0 represents background. However, the final image may contain some non-text part. Final result is the white text in black background or vice versa, dependent on the original image. Hence, text present is an image is well segmented from the background. Segmentation of the scene image into text and foreground is usually referred as binarization where grayscale intensities are classified into two groups, one is foreground white pixels and background black pixels (text) [13] [14]. The result is as shown in the figure: 7

Figure 7: Combination of each binarized plane and its complement version after connected component analysis.

E. Morphological operations and detection of text in image

Using morphological dilation operation we can localize text in the scene images. The dilat-ion of an image a by a SE b produce a new binary image k =a ⊕ b with ones in all locat-ions (x, y) of a SE origin at which that SE s hits the input image a , i.e. k(x, y) = 1 if b hits a .and 0 otherwise, repeating for all coordinates (x, y) [12]. Let a c denote the complement of an image a, i.e., the image produced .by replacing 1 with 0 and vice versa. Formally, the duality is written as:

$$a \oplus b = a^{c} \oplus b\_rot \qquad (2)$$

where brot is the SE b rotated by 〖 180〗 ^∘. If a SE is symmetrical with respect to rotation, then brot does not differ from b [8]. Morphological dilation is used for this purpose as dilation adds pixels to the boundaries. of objects in an image there by thickening that object. Measure thickness is defined by the type and size SE. Proper sized SE should be chosen such that least non-text area should be clustered within. Here, structuring element "line" with size (a line of degree 150) is used. The localized text region obtained is as shown below

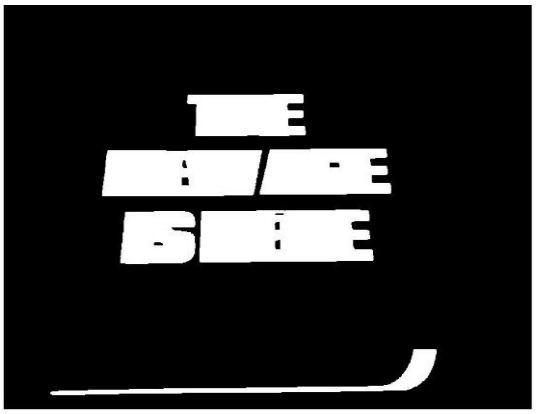


Figure 8: localization of text

After localizing text a bounding box is used to detect the text regions. Once we put bounding box we can easily extract text regions. As expected, many of the extracted connected components do not actually contain text characters. At this point simple rules are used to filter out the false detections. We use the aspect ratio and area size to decrease the number of non-character candidates [9]. Wi and Hi are the width and height of an extracted area; Δx and Δy are the distances between the centers of gravity of each area. Aspect ratio is computed as width / height. We use the following rules to further eliminate from all the detected connected components those that do not actually correspond to text characters.

$$0.1 < \frac{W_i}{H_i} < 20.5 < H_i / H_j < 2 \quad (3)$$

$$\Delta y < 0.2 \times \max(H_i, H_j) \quad (4)$$

$$\Delta x < 2 \times \max(W_i, W_j) \quad (5)$$

Selected C.Cs form the segmented text at the pixel level. The bounding box list provides the localisation of text in the given image. The detected text is placed in a green boundary. The result is shown in the Figure 8. Hence, opening of an object A with a linear structuring element B can effectively identify the horizontal line segments present in a connected component

Figure 9: Detected Text is placed in green boundary

IV. RESULTS AND DISCUSSIONS

For evaluating the performance of the proposed text detection method, we used the dataset made available with the occasion of KAIST [10] and SVT [11].As each image contains approximately four characters of different font styles of different font size. The result of the system for 500 images is given in the Table (1).

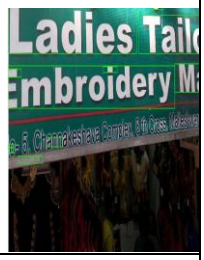
TABLE1: Summary of tested images.

DATASET	SEGMENTATION AND DETECTION		
	Number of Images	Segmented	Detected
KAIST	200	155	105
SVT	300	245	135

The experimental results is shown in Table 2. In the second row (second image) even though the text is well segmented from an image, the method is failed to detect all text regions due to reflection of light in the image. Hence, the method is restricted to extract text on glass surface. On rest of the lost images the algorithm either partially extracted relevant text components or extracted text along with a few non-text components. High accuracy result obtain if an image has clear background and normal font styles.

In summary, the precision and recall values of our algorithm obtained on the basis of the present set 500 images are respectively 69.8% and 71.2%. The proposed algorithm works well even on slanted or curved text components of English.

TABLE2: Result Images of Binarized Text and Detected text.

Original Image	Binarized Image	Text-region detection
		
		
		
		
		

CONCLUSION

The proposed method for text localization and segmentation the image, different from the known algorithms .Text localisation is required to segment and detect correctly. We have used SVT and KAIST dataset. Our method is tested for more than 400 scene images and detection rate 83.75% is achieved. Implemented algorithm has used morphological operations and connected component analysis using special domain features for extraction of text and detection of text in natural scene images. But due to diversity in scene images detection and segmentation performs well for the images with simple font style medium intensity variance, and simple background. Achieving higher accuracy, addressing complex background and light intensity variance will be the scope of the further research work.

REFERENCES

[1] Hongliang, Bai, and Liu Changping. "A hybrid license plate extraction method based on edge statistics and morphology." Pattern Recognition, ICPR. Proceedings of the 17th International Conference on. Vol. 2. IEEE, 2004

[2] Bai, Jinfeng, et al. "Chinese Image Character Recognition Using DNN and Machine Simulated Training Samples." Artificial Neural Networks and Machine Learning–ICANN Springer International Publishing, 2014. 209-216.

[3] Gupta, Neha, and V. K. Banga. "Image Segmentation for Text Extraction." Proceedings of the 2nd International Conference on Electrical, Electronics and Civil Engineering (ICEECE'2012), Singapore, April 28-29. 2012.

[4] Dutta, A., Pal, U., Bandyopadhya, A., & Tan, C. L. (2009). Gradient based Approach for Text Detection in Video Frames 1

[5] Sivasankaran, V., P. Chitra, and L. Roja. "Recognition of Text in Mobile Captured Images Based on Edge and Connected Component Hybrid Algorithm." International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE) 3.6 (2014): pp-358.

[6] Angadi, S. A., and M. M. Kodabagi. "Text region extraction from low resolution natural scene images using texture features." Advance Computing Conference (IACC), IEEE 2nd International. IEEE, 2010

[7] Wei, Yi Cheng, and Chang Hong Lin. "A robust video text detection approach using SVM." Expert Systems with Applications 39.12 (2012): 10832-10840.

[8] [8]. Xiaoqing Liu and Jagath Samarabandu, An Edge-based text region extraction algorithm for Indoor mobile robot navigation, Proceedings of the IEEE, July 2005.

[9] [9]. Xiaoqing Liu and Jagath Samarabandu, Multiscale edge-based Text extraction from Complex images, IEEE, 2006

[10] KAIST:http://www.iaprtc11.org/mediawiki/index.php/KAIST_Scene_Text_Database.

[11] SVT http://tc11.cvc.uab.es/datasets/SVT_1.